1. A group of researchers is running an experiment over the course of 30 months, with a single observation collected at the end of each month. Let $X_1, ..., X_{30}$ denote the observations for each month. From prior studies, the researchers know that

$$X_i \sim f_X(x),$$

but the mean μ_X is unknown, and they wish to conduct the following test

$$H_0 : \mu_X = 0$$
$$H_a : \mu_X > 0.$$

At month k, they have accumulated data $X_1, ..., X_k$ and they have the t-statistic

$$T_k = \frac{\bar{X} - 0}{S_k/\sqrt{n}}.$$

The initial plan was to test the hypotheses after all data was collected (at the end of month 30), at level $\alpha = 0.05$. However, conducting the experiment is expensive, so the researchers want to "peek" at the data at the end of month 20 to see if they can stop it early. That is, the researchers propose to check whether t_{20} provides statistically discernible support for the alternative. If it does, they will stop the experiment early and report support for the researcher's alternative hypothesis. If it does not, they will continue to month 30 and test whether t_{30} provides statistically discernible support for the alternative.

(a) What values of t_{20} provide statistically discernible support for the alternative hypothesis?

```
alpha = 0.05
m20 = 20
discernible_boundary_lower = qt(alpha/2, df = m20-1)
discernible_boundary_upper = qt(1-alpha/2, df = m20-1)

discernible_boundary_lower

## [1] -2.093024

discernible_boundary_upper

## [1] 2.093024
```

If $t_{20} > 2.093024$, then there is statistically discernible support for the alternative and we would reject the null.

(b) What values of t_{30} provide statistically discernible support for the alternative hypothesis?

```
alpha = 0.05
m30 = 30
boundary_lower = qt(alpha/2, df = m30-1)
boundary_upper = qt(1-alpha/2, df = m30-1)

boundary_lower

## [1] -2.04523

boundary_upper

## [1] 2.04523
```

If $t_{30} > 2.04523$, then there is statistically discernible support for the alternative and we would reject the null.

(c) Suppose $f_X(x)$ is a Laplace distribution with $a = 0$ and $b = 4.0$. Conduct a simulation study to assess the Type I error rate of this approach.
Note: You can use the `rlaplace()` function from the `VGAM` package for `R` (Yee, 2010).

```
library(VGAM)
#??rlaplace
alpha = 0.05
m20 = 20
m30 = 30
t_20 = qt(1-alpha, df = m20-1)
t_30 = qt(1-alpha, df = m30-1)

simu_type1 = function(N = 10000, b = 4){
  false_positives = 0
  for (i in 1:N){
    x = rlaplace(30, location = 0, scale = b)
    x20 = x[1:20]
    t20 = (mean(x20) - 0) / (sd(x20)/sqrt(20)) # Centered at 0 so mu = 0

    if (t20 > t_20){
      false_positives = false_positives + 1
    }
    else {
      t30 = (mean(x) - 0) / (sd(x)/sqrt(30))
      if (t30 > t_30){
        false_positives = false_positives + 1
      }
    }
  }
  return(false_positives / N)
}

type1_rate = simu_type1()
type1_rate

## [1] 0.0732
```

The Type I error rate for this approach is 0.0735 or 7.35%

(d) **Optional Challenge:** Can you find a value of $\alpha < 0.05$ that yields a Type I error rate of 0.05?

2. Perform a simulation study to assess the robustness of the T test. Specifically, generate samples of size $n = 15$ from the Beta(10,2), Beta(2,10), and Beta(10,10) distributions and conduct the following hypothesis tests against the actual mean for each case (e.g., $\frac{10}{10+2}$, $\frac{2}{10+2}$, and $\frac{10}{10+10}$).

   (a) What proportion of the time do we make an error of Type I for a left-tailed test?

   (b) What proportion of the time do we make an error of Type I for a right-tailed test?

   (c) What proportion of the time do we make an error of Type I for a two-tailed test?

   (d) How does skewness of the underlying population distribution effect Type I error across the test types?

```
t_test_errors = function(N = 10000, func, true_mean, n = 15){
  left_error = 0
  right_error = 0
  two_error = 0
```

```r
  for (i in 1:N){
    x = func(n)
    t_stat = (mean(x) - true_mean) / (sd(x) / sqrt(n))
    if (t_stat < qt(0.05, df = n - 1)){ # Checks the left-tailed test error rate
      left_error = left_error + 1
    }
    if (t_stat > qt(0.95, df = n - 1)){
      right_error = right_error + 1
    }
    if(abs(t_stat) > qt(0.975, df = n - 1)){
      two_error = two_error + 1
    }
  }
  return(c("Left-tailed" = left_error/N,
           "Right-tailed" = right_error/N,
           "Two-tailed" = two_error/N))
}
set.seed(1313)
beta_10_2 = t_test_errors(func = function(n) rbeta(n, 10,2), true_mean = 10/12)
beta_2_10 = t_test_errors(func = function(n) rbeta(n, 2, 10), true_mean = 2/12)
beta_10_10 = t_test_errors(func = function(n) rbeta(n, 10, 10), true_mean = 10/20)

beta_10_2

##  Left-tailed Right-tailed   Two-tailed
##       0.0300       0.0804       0.0619

beta_2_10

##  Left-tailed Right-tailed   Two-tailed
##       0.0831       0.0318       0.0624

beta_10_10

##  Left-tailed Right-tailed   Two-tailed
##       0.0506       0.0532       0.0518
```

(a) For $\text{Beta}(10, 2)$ distribution, the right-tailed test has a higher Type I error rate of 8.04%, while the left-tailed test has a lower error rate of 3%.

(b) The $\text{Beta}(2, 10)$ distribution has a right-tailed test Type I error rate of 3.18% whereas the left-tailed test has an 8.31% error rate.

(c) The $\text{Beta}(10, 10)$ distribution has near equal Type I error rates for the right-tailed and left-tailed tests at about 5%.

(d) The skewness of the distributions distort the Type I error rates, particularly for the one-sided tests. In contrast, two-tailed test error rates, which all remained relatively close to 6%, seem to produce the least distorted Type I error rates for various skewed distributions. This suggests that two-tailed tests may give more reliable Type I error rates.

References

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34.